Vector Transport Free Riemannian LBFGS for Optimization on Symmetric Positive Definite Matrix Manifolds

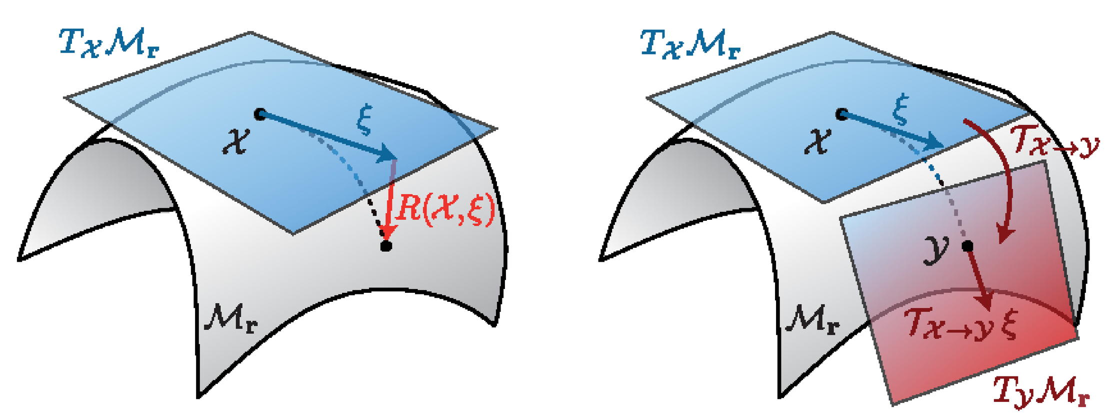
Reza Godaz[1] **Benyamin Ghojogh**[2] **Reshad Hosseini**[3] **Reza Monsefi**[1] **Fakhri Karray**[2] **Mark Crowley**[2]

[1]Ferdowsi University of Mashhad [2]University of Waterloo [3]University of Tehran

Riemannian Optimization

Optimization on a manifold where the point belongs to a manifold:

$$\underset{\Sigma}{\text{minimize}} \quad f(\Sigma) \qquad \text{subject to } \Sigma \in \mathcal{M}. \tag{1}$$



(a) Retraction (b) Vector transport

VTF-RLBFGS Algorithm

Algorithm 1: The VTF-RLBFGS algorithm

Input: Initial point $\boldsymbol{\Sigma}_0$

$\boldsymbol{H}_0 := \frac{1}{\sqrt{g'_{\boldsymbol{\Sigma}_0}(\nabla' f(\boldsymbol{\Sigma}_0), \nabla' f(\boldsymbol{\Sigma}_0))}} \boldsymbol{I}$

for $k = 0, 1, \dots$ **do**

- Compute $\nabla' f(\boldsymbol{\Sigma}_k)$ from Euclidean gradient by one of the mappings in Table 1
- $\boldsymbol{\xi}'_k :=$ GetDirection$(-\nabla' f(\boldsymbol{\Sigma}_k), k)$
- $\alpha_k :=$ Line search with Wolfe conditions
- $\boldsymbol{\Sigma}_{k+1} := \text{Exp}_{\boldsymbol{\Sigma}_k}(\alpha_k \boldsymbol{\xi}'_k)$ or $\text{Ret}_{\boldsymbol{\Sigma}_k}(\alpha_k \boldsymbol{\xi}'_k)$
- $\boldsymbol{s}'_{k+1} := \alpha_k \boldsymbol{\xi}'_k$
- $\boldsymbol{y}'_{k+1} := \nabla' f(\boldsymbol{\Sigma}_{k+1}) - \nabla' f(\boldsymbol{\Sigma}_k)$
- $\boldsymbol{H}_{k+1} := \frac{g'_{\boldsymbol{\Sigma}_{k+1}}(\boldsymbol{s}'_{k+1}, \boldsymbol{y}'_{k+1})}{g'_{\boldsymbol{\Sigma}_{k+1}}(\boldsymbol{y}'_{k+1}, \boldsymbol{y}'_{k+1})}$
- Store $\boldsymbol{y}'_{k+1}$, $\boldsymbol{s}'_{k+1}$, $g'_{\boldsymbol{\Sigma}_{k+1}}(\boldsymbol{s}'_{k+1}, \boldsymbol{y}'_{k+1})$, $g'_{\boldsymbol{\Sigma}_{k+1}}(\boldsymbol{s}'_{k+1}, \boldsymbol{s}'_{k+1})$, and $\boldsymbol{H}_{k+1}$

end

return $\boldsymbol{\Sigma}_{k+1}$

Function GetDirection$(\boldsymbol{p}', k)$

if $k > 0$ **then**

- $\rho_k := \frac{1}{g'_{\boldsymbol{\Sigma}_k}(\boldsymbol{y}'_k, \boldsymbol{s}'_k)}$
- $\widetilde{\boldsymbol{p}}' := \boldsymbol{p}' - \rho_k\, \boldsymbol{g}'_{\boldsymbol{\Sigma}_k}(\boldsymbol{s}'_k, \boldsymbol{p}') \boldsymbol{y}'_k$
- $\widehat{\boldsymbol{p}}' :=$ GetDirection$(\widetilde{\boldsymbol{p}}', k-1)$
- **return** $\widehat{\boldsymbol{p}}' - \rho_k\, \boldsymbol{g}'_{\boldsymbol{\Sigma}_k}(\boldsymbol{y}'_k, \widehat{\boldsymbol{p}}'_k)\boldsymbol{s}'_k + \rho_k\, \boldsymbol{g}'_{\boldsymbol{\Sigma}_k}(\boldsymbol{s}'_k, \boldsymbol{s}'_k)\boldsymbol{p}'$

else

- **return** $\boldsymbol{H}_0 \boldsymbol{p}'$

end

Mapping Operators on SPD Manifold

Operator	No mapping
Metric, $g_\Sigma(\boldsymbol{\xi}, \boldsymbol{\eta})$	$\textbf{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta})$
Gradient, $\nabla f(\Sigma)$	$\frac{1}{2}\boldsymbol{\Sigma}\big(\nabla_E f(\boldsymbol{\Sigma}) + (\nabla_E f(\boldsymbol{\Sigma}))^\top\big)\boldsymbol{\Sigma}$
Exponential map, $\text{Exp}_\Sigma(\boldsymbol{\xi})$	$\boldsymbol{\Sigma}\exp(\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}) = \boldsymbol{\Sigma}^{\frac{1}{2}}\exp(\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\xi}\boldsymbol{\Sigma}^{-\frac{1}{2}})\,\boldsymbol{\Sigma}^{\frac{1}{2}}$
Vector transport, $\mathcal{T}_{\Sigma_1,\Sigma_2}(\boldsymbol{\xi})$	$\boldsymbol{\Sigma}_2^{\frac{1}{2}}\boldsymbol{\Sigma}_1^{-\frac{1}{2}}\boldsymbol{\xi}\boldsymbol{\Sigma}_1^{-\frac{1}{2}}\boldsymbol{\Sigma}_2^{\frac{1}{2}}$ or $\boldsymbol{L}_2\boldsymbol{L}_1^{-1}\boldsymbol{\xi}\boldsymbol{L}_1^{-\top}\boldsymbol{L}_2^\top$
Approx. Euclidean retraction, $\text{Ret}_\Sigma(\boldsymbol{\xi})$	$\boldsymbol{\Sigma} + \boldsymbol{\xi} + \frac{1}{2}\boldsymbol{\xi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}$
Operator	**Mapping by inverse second root**
Mapping	$\boldsymbol{\xi}' := \boldsymbol{\Sigma}^{-\frac{1}{2}}\,\boldsymbol{\xi}\,\boldsymbol{\Sigma}^{-\frac{1}{2}}$
Metric, $g'_\Sigma(\boldsymbol{\xi}', \boldsymbol{\eta}')$	$\textbf{tr}(\boldsymbol{\xi}'\boldsymbol{\eta}')$
Gradient, $\nabla' f(\Sigma)$	$\frac{1}{2}\boldsymbol{\Sigma}^{\frac{1}{2}}\big(\nabla_E f(\boldsymbol{\Sigma}) + (\nabla_E f(\boldsymbol{\Sigma}))^\top\big)\boldsymbol{\Sigma}^{\frac{1}{2}}$
Exponential map, $\text{Exp}_\Sigma(\boldsymbol{\xi}')$	$\boldsymbol{\Sigma}^{\frac{1}{2}}\exp(\boldsymbol{\xi}')\,\boldsymbol{\Sigma}^{\frac{1}{2}}$
Vector transport, $\mathcal{T}'_{\Sigma_1,\Sigma_2}(\boldsymbol{\xi}')$	$\boldsymbol{\xi}'$
Approx. Euclidean retraction, $\text{Ret}_\Sigma(\boldsymbol{\xi}')$	$\boldsymbol{\Sigma} + \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\xi}'\boldsymbol{\Sigma}^{\frac{1}{2}} + \frac{1}{2}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\xi}'^2\boldsymbol{\Sigma}^{\frac{1}{2}}$
Operator	**Mapping by Cholesky decomposition**
Mapping	$\boldsymbol{\xi}' := \boldsymbol{L}^{-1}\,\boldsymbol{\xi}\,\boldsymbol{L}^{-\top}$
Metric, $g'_\Sigma(\boldsymbol{\xi}', \boldsymbol{\eta}')$	$\textbf{tr}(\boldsymbol{\xi}'\boldsymbol{\eta}')$
Gradient, $\nabla' f(\Sigma)$	$\frac{1}{2}\boldsymbol{L}^\top\big(\nabla_E f(\boldsymbol{\Sigma}) + (\nabla_E f(\boldsymbol{\Sigma}))^\top\big)\boldsymbol{L}$
Exponential map, $\text{Exp}_\Sigma(\boldsymbol{\xi}')$	$\boldsymbol{\Sigma}\exp(\boldsymbol{L}^{-\top}\boldsymbol{\xi}'\boldsymbol{L}^\top)$
Vector transport, $\mathcal{T}'_{\Sigma_1,\Sigma_2}(\boldsymbol{\xi}')$	$\boldsymbol{\xi}'$
Approx. Euclidean retraction, $\text{Ret}_\Sigma(\boldsymbol{\xi}')$	$\boldsymbol{\Sigma} + \boldsymbol{L}\,\boldsymbol{\xi}'\boldsymbol{L}^\top + \frac{1}{2}\boldsymbol{L}\,\boldsymbol{\xi}'^2\boldsymbol{L}^\top$

Riemannian LBFGS

Limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS):

$$\widetilde{\boldsymbol{p}} := \boldsymbol{p} - \rho_k\, \boldsymbol{g}_{\Sigma_k}(\boldsymbol{s}_k, \boldsymbol{p})\boldsymbol{y}_k, \tag{2}$$

$$\widehat{\boldsymbol{p}} := \mathcal{T}_{\Sigma_{k-1},\Sigma_k}\big(\text{GetDirection}(\mathcal{T}^*_{\Sigma_{k-1},\Sigma_k}(\widetilde{\boldsymbol{p}}), k-1)\big), \tag{3}$$

$$\text{return } \boldsymbol{\xi}_k := \widehat{\boldsymbol{p}} - \rho_k\, \boldsymbol{g}_{\Sigma_k}(\boldsymbol{y}_k, \widehat{\boldsymbol{p}})\boldsymbol{s}_k + \rho_k\, \boldsymbol{g}_{\Sigma_k}(\boldsymbol{s}_k, \boldsymbol{s}_k)\boldsymbol{p}, \tag{4}$$

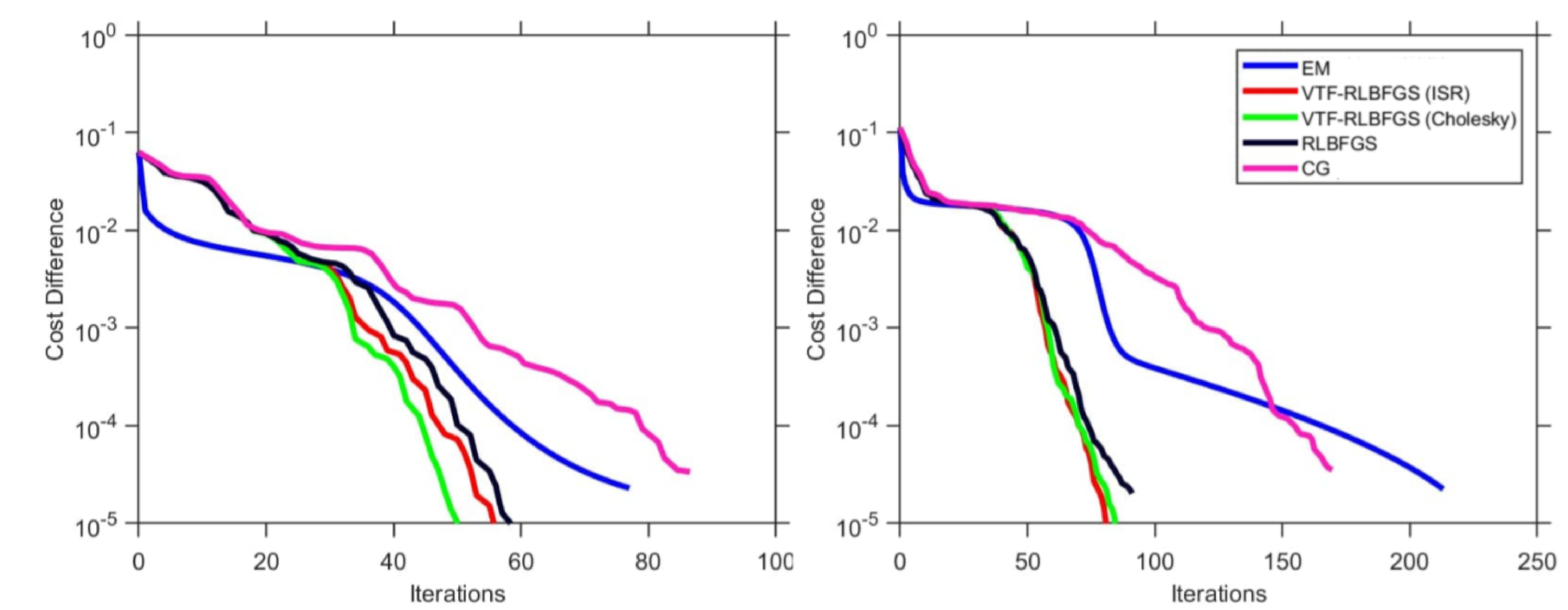
$$\Sigma_{k+1} := \text{Exp}_{\Sigma_k}(\alpha_k \boldsymbol{\xi}_k) \text{ or } \text{Ret}_{\Sigma_k}(\alpha_k \boldsymbol{\xi}_k), \tag{5}$$

$$\boldsymbol{s}_{k+1} := \mathcal{T}_{\Sigma_k,\Sigma_{k+1}}(\alpha_k \boldsymbol{\xi}_k), \tag{6}$$

$$\boldsymbol{y}_{k+1} := \nabla f(\Sigma_{k+1}) - \mathcal{T}_{\Sigma_k,\Sigma_{k+1}}\big(\nabla f(\Sigma_k)\big). \tag{7}$$

Experiments on Gaussian Mixture Model

$$\underset{\{\alpha_j,\boldsymbol{\mu}_j,\Sigma_j\}_{j=1}^K}{\text{minimize}} \; -\sum_{i=1}^N \log\Big(\sum_{j=1}^K \alpha_j\, \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{\mu}_j, \Sigma_j)\Big), \qquad \text{subject to } \; \Sigma_j \in \mathcal{M} = \mathbb{S}^n_{++}, \quad \forall j \in \{1, \dots, K\}, \tag{8}$$



(a) (b)

Experiments on Geometric Metric Learning

$$\min_{\boldsymbol{W}} \; f := \sum_{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\mathcal{S}} (\boldsymbol{x}_i - \boldsymbol{x}_j)^\top \boldsymbol{W}(\boldsymbol{x}_i - \boldsymbol{x}_j) + \sum_{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\mathcal{D}} (\boldsymbol{x}_i - \boldsymbol{x}_j)^\top \boldsymbol{W}^{-1}(\boldsymbol{x}_i - \boldsymbol{x}_j) + \frac{1}{2}\|\boldsymbol{W}\|_F^2, \qquad \text{s.t. } \; \boldsymbol{W} \in \mathcal{M} = \mathbb{S}^n_{++}, \tag{9}$$

Data	Algorithm	#iters	conv. time	iter. time	last cost
Iris	VTF (ISR)	23.500±4.528	2.461±1.174	0.110±0.070	1512.602±347.004
	VTF (Chol.)	25.000±5.676	2.474±1.009	0.096±0.028	1594.975±124.087
	RLBFGS	25.000±4.714	2.914±1.239	0.113±0.037	1620.207±125.989
USPS	VTF (ISR)	14.700±3.234	1.885±1.024	0.133±0.094	14223.215±0.001
	VTF (Chol.)	13.100±1.524	1.246±0.217	0.095±0.012	14223.216±0.001
	RLBFGS	13.100±2.234	1.307±0.454	0.098±0.025	14223.215±0.001
MNIST	VTF (ISR)	11.700±3.335	1.288±0.619	0.108±0.041	6254.283±0.001
	VTF (Chol.)	13.100±3.479	1.435±0.793	0.104±0.029	6254.284±0.001
	RLBFGS	12.100±3.381	1.266±0.754	0.099±0.024	6254.284±0.001